Stochastic Volatility in Mean Model: Empirical evidence from international stock markets

Koopman, S.J.; Hol Uspensky, E.

2000

document version
Early version, also known as pre-print

TI 2000-024/4
Tinbergen Institute Discussion Paper

The Stochastic Volatility in Mean Model

Siem Jan Koopman
Eugenie Hol Uspensky

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at
http://www.tinbergen.nl

The Stochastic Volatility in Mean model: Empirical evidence from international stock markets

Siem Jan Koopman and Eugenie Hol Uspensky
Department of Econometrics, Free University Amsterdam
Department of Accounting and Finance, University of Birmingham

November 11, 1999

Abstract

In this paper we present an exact maximum likelihood treatment for the estimation of a Stochastic Volatility in Mean (SVM) model based on Monte Carlo simulation methods. The SVM model incorporates the unobserved volatility as an explanatory variable in the mean equation. The same extension is developed elsewhere for Autoregressive Conditional Heteroskedastic (ARCH) models, known as the ARCH in Mean (ARCH-M) model. The estimation of ARCH models is relatively easy compared with that of the Stochastic Volatility (SV) model. However, efficient Monte Carlo simulation methods for SV models have been developed to overcome some of these problems. The details of modifications required for estimating the volatility-in-mean effect are presented in this paper together with a Monte Carlo study to investigate the small-sample properties of the SVM estimators. Taking these developments of estimation methods into account, we regard SV and SVM models as practical alternatives to their ARCH counterparts and therefore it is of interest to study and compare the two classes of volatility models. We present an empirical study about the intertemporal relationship between stock index returns and their volatility for the United Kingdom, United States and Japan. This phenomenon has been discussed in the financial literature but has proved hard to find empirically; we find evidence of a negative but weak relationship.

KEYWORDS: Forecasting, GARCH, Simulated maximum likelihood, Stochastic volatility, Stock indices.

1 Introduction

It is generally acknowledged that the volatility of many financial return series is not constant over time and that these series exhibit prolonged periods of high and low volatility, often referred to as volatility clustering. Over the past two decades two prominent classes of models have been developed which capture this time-varying autocorrelated volatility process: the Generalised Autoregressive Conditional Heteroscedasticity (GARCH) and the Stochastic Volatility (SV) model. GARCH models define the time-varying variance as a deterministic function of past squared innovations and lagged conditional variances whereas the variance in the SV model is modelled as an unobserved component that follows some stochastic process[1]. The most popular version of the SV model defines volatility as a logarithmic first order autoregressive process, which is a discrete-time approximation of the continuous-time Ornstein-Uhlenbeck diffusion process used in the option pricing literature[2].

[1]For surveys on the extensive GARCH literature we refer to Bollerslev, Chou and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle and Nelson (1994) and Diebold and Lopez (1995). SV models are reviewed in, for example, Taylor (1994), Ghysels, Harvey and Renault (1996) and Shephard (1996).

[2]See Hull and White (1987), Scott (1987) and Wiggins (1987) and Chesney and Scott (1989).

Although SV models are seen as a competitive alternative to GARCH models their empirical application has been limited. This can mainly be attributed to the difficulties that arise as a result of the intractability of the likelihood function which prohibits its direct evaluation. However, in recent years considerable advances have been made in this area. The estimation techniques that have been proposed for SV models can be divided into two groups: those that seek to construct the full likelihood function and those that approximate it or avoid the issue altogether. The methods originally suggested by Taylor (1986) and Harvey, Ruiz and Shephard (1994) belong to the latter category. Recently attention has moved towards the development of techniques that attempt to evaluate the full likelihood function[3]. For recent reviews on these full likelihood methods we refer to Sandmann and Koopman (1998) and Fridman and Harris (1998). The estimation method we adopt here is based on the Monte Carlo likelihood approach developed by Shephard and Pitt (1997) and Durbin and Koopman (1997) where the likelihood function is evaluated using importance sampling. These new techniques enable us to include explanatory variables in the mean equation and estimate their coefficients simultaneously with the parameters of the volatility process[4]. One of the explanatory variables in our model is the variance process itself, hence its name: Stochastic Volatility in Mean (SVM). The estimation of such an intricate model is not straightforward since volatility now appears in both the mean and the variance equation. This requires modification of the simulation maximum likelihood estimation method, details of which are given in section 3.

The SV models we present are a practical alternative to the GARCH type models that have been used so widely in empirical financial research and which have relied on simultaneous modeling of the first and second moment. For certain financial time series such as stock index returns, which have been shown to display high positive first order autocorrelations, this constitutes an improvement in terms of efficiency; see Campbell, Lo and MacKinlay (Chapter 2, 1997). The volatility of daily stock index returns has been estimated with SV models but usually results have relied on extensive pre-modelling of these series, thus avoiding the problem of simultaneous estimation of the mean and variance[5]. The fact that we are able to estimate an SV model that includes volatility as one of the determinants of the mean makes our model suitable for empirical applications in which returns are partially dependent on volatility, such as studies that investigate the relationship between the mean and variance of stock returns. The SVM model can therefore be viewed as the SV counterpart of the ARCH-M model of Engle, Lilien and Robins (1987). In section 4 we investigate the intertemporal relationship between daily stock index returns and their volatility for three international stock indices and compare the resulting parameter estimates with those obtained for GARCH-M models. The stock indices we examine are the Financial Times All Share (UK), the Standards & Poor Composite (US) and the Topix Index (Japan).

The remainder of this paper is organised as follows. The specification of time-varying variance models in general and the SVM model in particular are discussed in section 2. In section 3 we develop the simulated maximum likelihood estimation method for the SVM model. Further, some Monte Carlo evidence of small sample consistency of the estimated parameters is given. Section 4 describes the stock index data and reports on parameter estimation results. In the final section we present a summary and some conclusions.

---

[3] See, for example, Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998), Sandmann and Koopman (1998) and Fridman and Harris (1998).

[4] Also see Fridman and Harris (1998) and Chib, Nadari and Shephard (1998).

[5] The same seasonally adjusted S&P Composite stock index series (Gallant, Rossi and Tauchen, 1992) has been used in a number of studies, see for example: Jacquier *et. al.* (1994), Danielsson (1994), Sandmann and Koopman (1998), Fridman and Harris (1998) and Chib *et.al.* (1998).

2 Modelling Volatility

2.1 Basic model

The aim is to simultaneously model the mean and variance of a series of returns on an asset denoted by y_t. Both the SV and GARCH model are defined by their first and second moment which can be referred to as the mean and variance equation. The most general form of the mean equation for both models is then defined as

$$y_t = \mu_t + \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0,1), \tag{1}$$

$$\mu_t = a + \sum_{i=1}^{k} b_i x_{i,t}, \tag{2}$$

where μ_t denotes the conditional mean which depends on a constant a and regression coefficients $b_1, \ldots, b_k$. The explanatory variables $x_{i,t}$ may also contain lagged exogeneous and dependent variables. The disturbance term ε_t is independently and identically distributed with zero mean and unit variance. Usually, the assumption of a normal distribution for ε_t is added. The positive volatility process is denoted by σ_t which remains to be specified in section 2.2 for GARCH and section 2.3 for SV models. The mean adjusted series is therefore defined as white noise multiplied by the volatility process.

2.2 GARCH model

The general form of the GARCH(p,q) model is

$$\begin{aligned} \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i (y_{t-i} - \mu_{t-i})^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \\ &= \omega + \sum_{i=1}^{p} \alpha_i (\sigma_{t-i} \varepsilon_{t-i})^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \end{aligned} \tag{3}$$

where the parameters to be estimated are ω, $\alpha_1, \ldots, \alpha_p$ and $\beta_1, \ldots \beta_q$. An unanticipated shock to the return process at time t is therefore not incorporated into the volatility process until time $t+1$.

The most commonly used model in applied financial studies is the GARCH(1,1) model which is given by

$$\sigma_t^2 = \omega + \alpha (y_{t-1} - \mu_{t-1})^2 + \beta \sigma_{t-1}^2, \tag{4}$$

with parameter values restricted to $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$. Provided that the sum of α and β is less than one, the unconditional expectation of the conditional variance is constant and finite and given by

$$\frac{\omega}{1 - \alpha - \beta}.$$

In empirical financial research with high frequency data, $\alpha + \beta$ is often estimated as being close to unity, which implies a high degree of volatility persistence. Apart from volatility clustering GARCH models also capture part of the excess kurtosis observed in financial time series. Under the assumption of normality, existence of the fourth order moment for the GARCH(1,1) model is ensured if $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$. Subject to this restriction it can be shown that the fourth moment will exhibit excess kurtosis

$$\kappa_y = \frac{\kappa_\varepsilon \, \text{E}(\sigma_t^4)}{\text{E}(\sigma_t^2)^2} = 3 + \frac{6\alpha^2}{1 - \beta^2 - 2\alpha\beta - 3\alpha^2},$$

and therefore $\kappa_y > \kappa_\varepsilon$; see Bollerslev (1986). For a further discussion on the features of GARCH models we refer to a number of surveys such as the ones given in note 1.

2.3 SV model

In the case of the SV model the variance equation is specified in logarithmic form, that is

$$\sigma_t = \sigma^* \exp(0.5 h_t) \tag{5}$$

with positive scaling factor σ^*. It follows that $h_t = \ln(\sigma_t^2/\sigma^{*2})$ where the stochastic process for h_t is

$$h_t = \phi h_{t-1} + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathrm{NID}(0,1), \tag{6}$$

with persistence parameter ϕ which is restricted to a positive value less than one to ensure stationarity. The disturbances ε_t and η_t are mutually uncorrelated, contemporaneously and at all lags. The unconditional variance implied by the SV model is given by

$$\sigma^{*2} \exp\left(0.5 \frac{\sigma_\eta^2}{1-\phi^2}\right),$$

and it can be shown that this model also captures part of the excess kurtosis as

$$\kappa_y = \frac{\kappa_\varepsilon \, \mathrm{E}(\sigma_t^4)}{\mathrm{E}(\sigma_t^2)^2} = 3 \exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right),$$

which also implies that $\kappa_y > \kappa_\varepsilon$. Alternative specifications for the SV model can be deduced from

$$\begin{aligned}
\ln \sigma_t^2 &= \ln \sigma^{*2} + h_t \\
&= \ln \sigma^{*2} + \phi(\ln \sigma_{t-1}^2 - \ln \sigma^{*2}) + \sigma_\eta \eta_t \\
&= (1-\phi) \ln \sigma^{*2} + \phi \ln \sigma_{t-1}^2 + \sigma_\eta \eta_t.
\end{aligned}$$

The main distinction between GARCH and SV models is that the latter has seperate disturbance terms in the mean and variance equation, ε_t and η_t, respectively, which precludes direct observation of the variance process σ_t^2. GARCH models are deterministic in the sense that only the mean equation has a disturbance term and that its variance is modelled conditionally on I_{t-1}, that is the information upto and including time $t-1$. Therefore, the variance can be observed at time t. For the SV model, the deviation of y_t from the mean is captured by a function of the two disturbance terms whereas in the GARCH model this deviation is accounted for by a single disturbance term. For the GARCH model this point is evident but to clearify this for the SV model, we rewrite the model as follows:

$$\begin{aligned}
y_t &= \mu_t + \sigma_t \varepsilon_t \\
&= \mu_t + \sigma^* \exp(0.5 h_t) \varepsilon_t \\
&= \mu_t + \sigma^* \exp(0.5 \phi h_{t-1}) \exp(0.5 \eta_t) \varepsilon_t.
\end{aligned}$$

The overall innovation term of the SV model is the error term $\exp(0.5\eta_t)\varepsilon_t$ with a zero mean but with a non-Gaussian density.

2.4 Volatility in mean

The SV model with volatility included in the mean is given by (1) and (5) where the mean equation (2) is rewritten as

$$\mu_t = a + \sum_{i=1}^{k} b_i x_{i,t} + d\sigma^{*2} \exp(h_t), \tag{7}$$

with d as the regression coefficient measuring the volatility-in-mean effect. In particular, we will use the mean specification

$$\mu_t = a + b y_{t-1} + d\sigma^{*2} \exp(h_t). \tag{8}$$

This SVM model has six parameters which are to be estimated simultaneously using simulation methods which will be discussed in the next section. Inclusion of the variance as one of the determinants of the mean facilitates the examination of the relationship between returns and volatility. It enables us to perform studies in the vein of French, Schwert and Stambaugh (1987) but in the context of SV models. The relative ease with which they were able to conduct their research, *i.e.* without prior manipulation of the original data series, is now also feasible for SV models.

The equivalent in mean specification for the GARCH model is

$$\mu_t = a + by_{t-1} + d\sigma_t^2. \tag{9}$$

3 Estimation of the SVM model

In this section we show how the parameters of the SVM model are estimated by simulated maximum likelihood. Further, we show how to compute the conditional mean and variance of the volatility process h_t.

3.1 Model

To simplify the exposition we initially consider the model

$$\begin{aligned} y_t &= d\sigma^{*2}\exp(h_t) + \sigma^*\exp(0.5h_t)\varepsilon_t, \\ h_t &= \phi h_{t-1} + \sigma_\eta \eta_t, \end{aligned}$$

where y_t denotes the underlying series of interest, in our case these are stock index returns. The details for the estimation of the full model will be given in section 3.7. The disturbances ε_t and η_t are standard normally distributed and mutually and serially uncorrelated. The latent variable h_t is modelled as a stationary Gaussian autoregressive process of order 1 and with $0 < \phi < 1$. The unknown parameters are collected in the vector

$$\psi = (\phi, \sigma_\eta, \sigma^{*2}, d)'.$$

The nature of the model is Gaussian but we deal with a nonlinear model since the variance of the overall disturbance term in y_t is given by $\sigma^{*2}\exp(h_t)$ which is stochastic. The Gaussian density for ε_t can be replaced by other continuous distributions.

Other formulations of the SVM model are possible but we have chosen this one since it is closely associated with the ARCH in Mean models; see section 2. From a technical point, the conditional density function $p(y|\theta, \psi)$ of the SVM model with

$$\theta = (h_1, \ldots, h_T)',$$

is log-concave in h_t. This property is required for the techniques used in the following sections.

3.2 Likelihood evaluation using importance sampling

The construction of the likelihood for the SVM model is complicated because the latent variable h_t appears in both the mean and the variance of the SVM model. We adopt the Monte Carlo likelihood approach developed by Shephard and Pitt (1997) and Durbin and Koopman (1997). This simulation method of computing the loglikelihood function can be derived as follows.

Define the likelihood as

$$L(\psi) = p(y|\psi) = \int p(y, \theta|\psi)d\theta = \int p(y|\theta, \psi)p(\theta|\psi)d\theta. \tag{10}$$

An efficient way of evaluating this likelihood is by using importance sampling; see Ripley (1987, chapter 5). We require a simulation device to sample from some importance density $\tilde{p}(\theta|y,\psi)$ which must be as close as possible to the true densitity $p(\theta|y,\psi)$. An obvious choice for the importance density is the conditional Gaussian density since in this case it is relatively straightforward to sample from $\tilde{p}(\theta|y,\psi) = g(\theta|y,\psi)$. In the Appendix an approximating Gaussian model for the SVM model is developed. The simulation smoother of de Jong and Shephard (1995) can be used to sample from the approximating Gaussian model $g(\theta|y,\psi)$.

The likelihood function (10) is rewritten as

$$L(\psi) = \int p(y|\theta,\psi)\frac{p(\theta|\psi)}{g(\theta|y,\psi)}g(\theta|y,\psi)d\theta = \tilde{\mathrm{E}}\{p(y|\theta,\psi)\frac{p(\theta|\psi)}{g(\theta|y,\psi)}\}, \tag{11}$$

where $\tilde{\mathrm{E}}$ denotes expectation with respect to the importance density $g(\theta|y,\psi)$. Expression (11) can be simplified considerably following a suggestion of Durbin and Koopman (1997). The likelihood function of the approximating Gaussian model is given by

$$L_g(\psi) = g(y|\psi) = \frac{g(y,\theta|\psi)}{g(\theta|y,\psi)} = \frac{g(y|\theta,\psi)p(\theta|\psi)}{g(\theta|y,\psi)}, \tag{12}$$

and it follows that

$$\frac{p(\theta|\psi)}{g(\theta|y,\psi)} = \frac{L_g(\psi)}{g(y|\theta,\psi)}.$$

This ratio also appears in (11) and substitution leads to

$$L(\psi) = L_g(\psi)\,\tilde{\mathrm{E}}\{\frac{p(y|\theta,\psi)}{g(y|\theta,\psi)}\}, \tag{13}$$

which is the convenient expression we will use in our calculations. The likelihood function of the approximating Gaussian model can be calculated via the Kalman filter and the two conditional densities are easy to compute given a value for θ. It follows that the likelihood function of the SVM model is equivalent to the likelihood function of an approximating Gaussian model, multiplied by a correction term. This correction term only needs to be evaluated via simulation.

An obvious estimator for the likelihood of the SVM model is

$$\hat{L}(\psi) = L_g(\psi)\bar{w}, \tag{14}$$

where

$$\bar{w} = \frac{1}{M}\sum_{i=1}^{M} w_i, \qquad w_i = \frac{p(y|\theta^i,\psi)}{g(y|\theta^i,\psi)}, \tag{15}$$

and θ^i denotes a draw from the importance density $g(\theta|y,\psi)$. The accuracy of this estimator solely depends on M, that is the number of simulation samples. In practice, we usually work with the log of the likelihood function to manage the magnitude of density values. The log transformation of $\hat{L}(\psi)$ introduces bias for which we can correct up to order $O(M^{-3/2})$; see Shephard and Pitt (1997) and Durbin and Koopman (1997). We obtain

$$\log\hat{L}(\psi) = \log L_g(\psi) + \log\bar{w} + \frac{s_w^2}{2M\bar{w}^2}, \tag{16}$$

with $s_w^2 = (M-1)^{-1}\sum_{i=1}^{M}(w_i-\bar{w})^2$.

3.3 Computational details

Given a particular vector for ψ, we evaluate the loglikelihood function (16) for which we use the approximating model (19) to generate simulation samples. To obtain a maximum likelihood estimate for ψ, which we denote by $\hat{\psi}$, the loglikelihood is numerically maximised with respect to ψ in a similar fashion as for Gaussian models; see Harvey (1989) and Koopman *et.al* (1995). The repeated evaluation of the loglikelihood for different ψ's during the search for $\hat{\psi}$ will be based on the same set of random numbers used for simulation.

Although the approximating model is effective for simulation, we may wish to decrease the simulation variance further using standard simulation techniques based on antithetics and control variables; see Durbin and Koopman (1997). In our computations we have only employed two antithetic variables. The first is the standard one given by $\breve{\theta}^i = 2\hat{\theta} - \theta^i$ where θ^i is a draw from the importance density $g(\theta|y,\psi)$ and where $\hat{\theta} = \tilde{\mathrm{E}}(\theta)$ can be obtained using the Kalman filter and smoother. Since $\breve{\theta}^i - \hat{\theta} = -(\theta^i - \hat{\theta})$ and θ^i are normally distributed, the two vectors θ^i and $\breve{\theta}^i$ are equi-probable. The second antithetic variable is proposed by Durbin and Koopman (1997) and it deals with balancing the variance within the generated simulation samples.

The number of simulation samples M is set prior to the estimation procedure. The choice of M can be determined by computing the error variance due to simulation; see Durbin and Koopman (1997). It is shown by Sandmann and Koopman (1998) that M can be relatively small in the context of SV models. Therefore, in this study we have set M equal to 50 times four antithetic variables, that is $M = 200$.

3.4 Signal extraction

The Monte Carlo importance sampling techniques, which we have used for likelihood evaluation, can also be employed to compute the conditional mean and variance of the unoberved process h_t. The same approximating Gaussian model can be used for this purpose. The details are given by Durbin and Koopman (2000).

Computation of the conditional mean and variance amounts to computing

$$\bar{h}_t^{(1)} = \frac{1}{M}\sum_{i=1}^{M} w_i h_t^i, \qquad \bar{h}_t^{(2)} = \frac{1}{M}\sum_{i=1}^{M} w_i h_t^{i2},$$

where w_i is defined in (15) and h_t^i is the tth element of θ^i which is the ith draw from the importance density $g(\theta|y,\psi)$. The conditional mean and variance of h_t is given by

$$\mathrm{E}(h_t|y,\psi) = \bar{h}_t^{(1)}, \qquad \mathrm{Var}(h_t|y,\psi) = \bar{h}_t^{(2)} - [\bar{h}_t^{(1)}]^2.$$

In practice, the unknown parameter vector ψ is replaced by its Monte Carlo maximum likelihood estimate $\hat{\psi}$. The uncertainty related to the estimate $\hat{\psi}$ can be also taken into account by similar Monte Carlo simulation techniques; see Durbin and Koopman (2000). An alternative approach of signal extraction for the SV model would be to adopt a Markov chain Monte Carlo techniques; see, for example, Shephard and Pitt (1997) and Kim, Shephard and Chib (1998).

3.5 Numerical implementation of estimation procedure

The simulated Monte Carlo estimation procedure is implemented using the object-oriented matrix programming language Ox 2.1 of Doornik (1998)[6] using the library SsfPack 2.3 of Koopman, Shephard and Doornik (1999)[7]. The relevant programs, including the one used for the Monte Carlo study in the

[6] More information at `www.nuffield.ox.ac.uk/users/Doornik/`
[7] More information at `www.ssfpack.com`

next subsection and the one used for the empirical study in section 4, can be downloaded from the Internet at `www.econ.vu.nl/koopman/sv/`. The programs can be adjusted in order to use them in a more general context (for example, with the inclusion of explanatory variables) and for other Monte Carlo studies. In addition, they can be applied to other data-sets. Documentation of the programs is available and can be consulted on-line.

3.6 Monte Carlo evidence of estimation procedure

In this section we present some results of a Monte Carlo study which is carried out to investigate the small sample performance of the estimation procedure presented in section 3.2. In short, we generate K simulated SVM series for the model presented in section 3.1 and for some given 'true' parameter vector ψ. Subsequently, we treat ψ as unknown and estimate it for each series using the maximum likelihood method described in section 3.2. We compute the sample mean and standard deviation together with a histogram for each element in ψ and compare it with the 'true' parameter value.

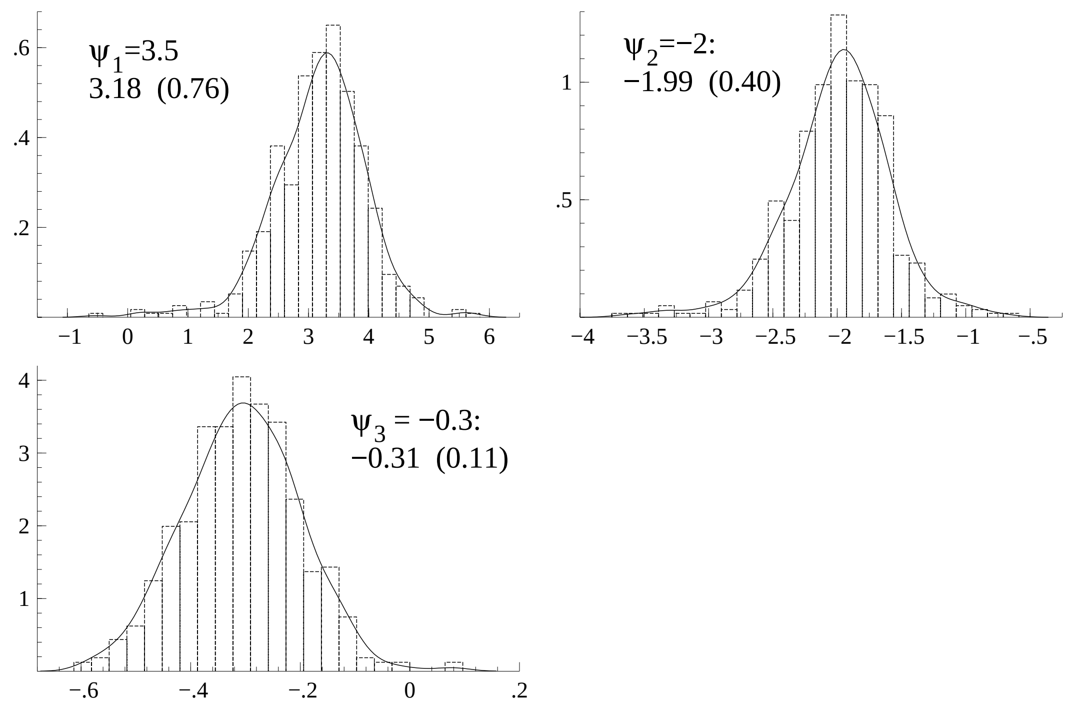

Figure 1: Monte Carlo results for standard SV model.
In each panel a histogram is presented and the 'true' parameter value for ψ_i^* is printed together with the mean and standard deviation (in parentheses) for the $K = 500$ maximum likelihood estimates for ψ_i^* for $i = 1, \ldots, 3$.

The estimation procedure is not with respect to ψ defined in section 3.1, but with respect to a transformed parameter vector ψ^*. The autoregressive parameter ϕ is restricted to have a value between zero and one; therefore we estimate ψ_1^* where

$$\phi = \psi_1 = \frac{\exp(\psi_1^*)}{1 + \exp(\psi_1^*)}, \qquad \psi_1^* = \log \frac{\phi}{1 - \phi}.$$

Further, we estimate the log variance σ^{*2} and the log standard deviation σ_η. The mean parameter d is estimated without transformation.

In the simulation exercises we have carried out we found satisfactory results. First, we considered the standard SV model. In this case, the last element of ψ does not play a role. For generating Monte Carlo samples, the remaining 'true' parameter values are set to

	ψ	ψ^*
$\psi_1 = \phi$	0.97	3.5
$\psi_2 = \sigma_\eta$	.135	-2
$\psi_3 = \sigma^{*2}$	.549	-0.3

which are typical values found in our empirical study of section 4.

The Monte Carlo results for the basic SV model are similar but slightly better compared to results presented in similar studies of Jacquier et al. (1994) and Sandmann and Koopman (1998). Note that in these studies the parameter values were not transformed and that the estimation procedures used were different from ours. For sample size $n = 500$ and the number of simulations set to $K = 500$, the results are given in figure 1. To present these results in terms of vector ψ, we note that the resulting confidence intervals are asymmetric due to the nonlinear transformations. We obtain

	mean	LHS "95% CI"	RHS "95% CI"
$\psi_1 = \phi = .97$	.957	.825	.990
$\psi_2 = \sigma_\eta = .135$	.139	.064	.302
$\psi_3 = \sigma^{*2} = .549$	.539	.351	.829

where LHS is the lefthand side border and RHS is the righthand side border of the 95% confidence interval. These results will be used as a benchmark for the Monte Carlo results for the SVM model.

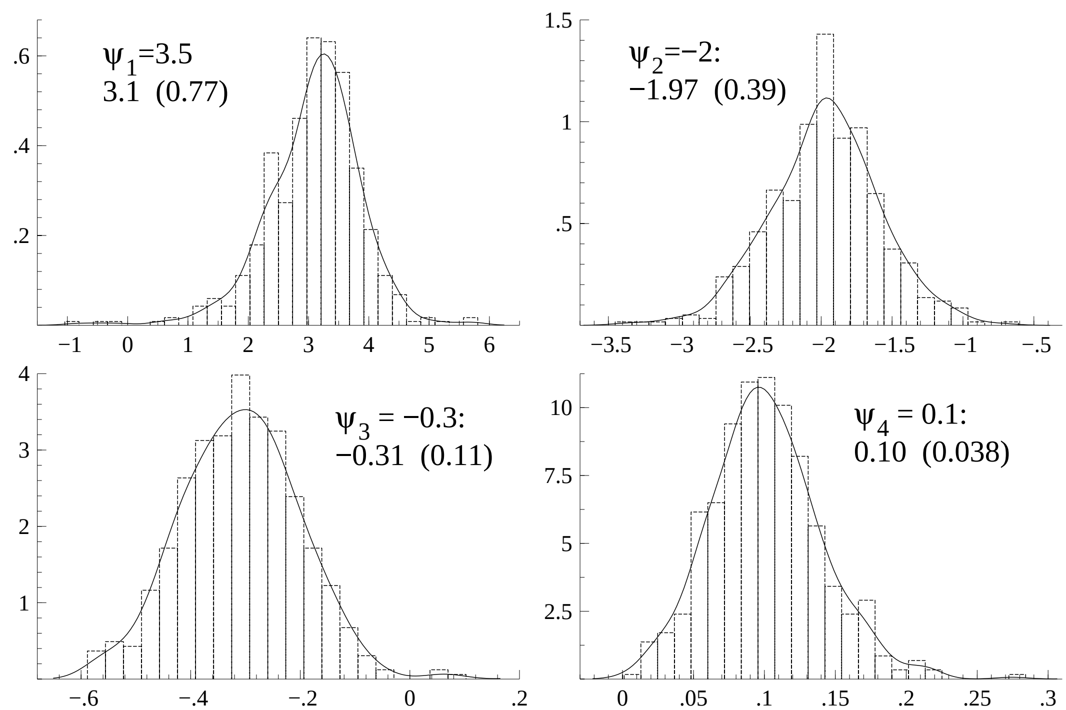

Figure 2: Monte Carlo results for the SVM model with $d = 0.1$.

We now turn our attention towards the Monte Carlo evidence for the SVM model. We keep the 'true' parameters of the SV model and look at the results for different values of d. The Monte Carlo experiments are again based on $n = 500$ and $K = 500$. The results for the SVM model with 'true' $d = 0.1$ are given in figure 2. The results in terms of ψ are given by

	mean	LHS "95% CI"	RHS "95% CI"
$\psi_1 = \phi = .97$	.957	.825	.990
$\psi_2 = \sigma_\eta = .135$	.139	.064	.302
$\psi_3 = \sigma^{*2} = .549$	.539	.351	.829
$\psi_4 = d = .1$	.101	.024	.178

Comparing the results for the standard SV model, we conclude that the confidence intervals for ψ_1, ψ_2 and ψ_3 are very similar. The in-mean parameter d is accurately estimated with a relatively small standard deviation. We repeated this experiment for the SVM model with various values of d and obtained similar results. As far as d is concerned, the estimates and corresponding standard deviations for different values of d are given by

'true'	mean	stand.dev
−0.10	−0.1009	0.0379
−0.05	−0.0492	0.0341
0.00	0.0016	0.0261
0.05	0.0503	0.0334
0.10	0.1014	0.0385

On the basis of the Monte Carlo evidence presented here, we conclude that the in-mean effect can be estimated accurately using the methods of section 3.2.

3.7 Extending the mean equation

The SVM model we have considered in section 3.1 can be extended to include a constant and a lagged dependent variable in the observation equation. We then obtain model (1) with σ_t given by equations (5) and (6) and μ_t given by (8). These extensions do not alter the estimation procedure as set out in section 3.2 because the state includes components which drive the variance σ_t^2. The extensions only influence the likelihood function via the squared error term. In the appendix, the last term of the definition for p_t is replaced by the term

$$\exp(-h_t)\sigma^{-2}\{y_t - a - by_{t-1} - d\exp(h_t)\}^2.$$

The extensions do not change the stochastic process for h_t. Therefore, the simulation scheme for computing the Monte Carlo likelihood remains the same.

The only real difference caused by the extension is that numerical maximisation of the Monte Carlo likelihood is also with respect to parameters a and b. Further, the approximating model, as derived in the appendix, changes slightly; that is, $\dot{p}_t$ changes but $\ddot{p}_t$ does not change. In other words, the definition for c_t changes but the definition for H_t does not.

Using the same arguments, we can include other explanatory variables in the observation equation. This implies that regression models with stochastic heteroskedasticity can be estimated using the techniques presented in this section. For example, we may consider the regression model

$$y_t = x_t'\delta + \sigma_t\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0,1),$$

for $t = 1, \ldots, n$ where x_t is a vector of explanatory variables, δ is a vector of coefficients and $\sigma_t^2 = \exp(h_t)$. The unobserved stochastic process h_t can be modelled within the state space form which allows for a wide range of different specifications. Such models may also be of interest outside the field of financial econometrics. Here we treat this subject as being beyond the scope of this paper.

4 Empirical evidence from international stock markets

4.1 Some theory on the relationship between returns and volatility

The intertemporal relationship between expected returns on stock indices and volatility has been the subject of a large number of studies. Many of these have used GARCH in Mean models to empirically investigate the validity of the Capital Asset Pricing Model (CAPM) which in essence is a single period equilibrium model without time dimension that defines expected excess returns on a market portfolio as a linear function of volatility so that

$$\mathrm{E}(y) = d\sigma^2, \tag{17}$$

where $E(y)$ is the expected differential between the return on a stock market portfolio and the risk-free rate of return, σ^2 denotes its variance and d is simply the ratio of the expected excess returns $E(y)$ and the variance σ^2.

The general mean equation with time-varying variance we consider for estimation is

$$\mu_t = a + by_{t-1} + d\sigma_t^2, \tag{18}$$

where σ_t^2 denotes the variance of y_t and the relation between excess returns and volatility is measured by d. In comparison with equation (17) two additional parameters are added: the constant in the mean term a and the autoregressive b parameter, where the latter is included to account for the first order autocorrelation customarily found in stock index return series.

A positive relationship, *i.e.* a positive value for the in-mean parameter d, appears plausible as rational risk-averse investors would require higher expected returns during more volatile periods when payoffs associated with these securities are less certain. Results reported in the GARCH literature are however inconclusive and it seems difficult to find evidence of a non-zero relationship. French, Schwert and Stambaugh (1987) and Campbell and Hentschel (1992) find evidence of a positive association, whereas Glosten, Jagannathan and Runkle (1993) who develop a much richer GARCH-M model observe a negative intertemporal relation for the US stock market, as does Nelson (1991) with his EGARCH model. Poon and Taylor (1992) who study the issue in a UK context report a positive yet weak relationship. These conflicting findings are however not without theoretical foundation: a positive relationship between expected returns and volatility over a given period is certainly persuasive but there is no consensus that, as Glosten *et.al.* (1993) point out, this relationship continues to hold across time and that it will be positive on average. The two variables are of course intimately related as increases in volatility are caused by large returns of either sign. What is more, empirical studies of asset returns have almost consistently shown that large negative returns occur more frequently than large positive ones, as one of the salient features of these asset return distributions is that they are negatively skewed. However, a negative value for the relationship does not automatically imply that the CAPM model is invalid as this one-period asset pricing model was never intended to explain the interdependence between contemporaneous expected returns and time-varying volatility.

The remainder of this section is organised as follows. We start by discussing the data of the three international stock market indices we selected in order to investigate the intertemporal relationship between excess returns and their volatility: the Financial Times All Share, the Standard & Poor's Composite and the Topix Index. We then proceed with the estimation results for these series using our SVM model. The parameter estimates are then compared with those obtained by the GARCH-M model. We also present results for alternative model specifications which we obtain by imposing restrictions on the various parameters in the mean.

4.2 Data

The data we analyse includes daily stock index returns from three international stock markets: the United Kingdom, the United States and Japan. The UK Financial Times All Share Index and the US

Table 1: Summary statistics of daily excess returns

Period	1975-1998		1988-1998		
Number of observations T	6261		2869		
Stock index	FT All	S&P	FT All	S&P	Topix
Mean	0.033	0.028	0.017	0.042	-0.025
Variance	0.943	0.874	0.584	0.747	1.357
Skewness	-0.194	-2.562	-0.022	-0.664	0.343
Excess Kurtosis	11.828	62.758	3.491	7.954	6.107
Excess Returns					
$\hat{\rho}_1$	0.167	0.054	0.115	0.004	0.100
$\hat{\rho}_2$	0.008	-0.024	-0.002	-0.013	-0.062
$\hat{\rho}_3$	0.037	-0.021	-0.005	-0.041	-0.009
$\hat{\rho}_4$	0.046	-0.024	0.041	-0.016	0.027
$\hat{\rho}_5$	0.019	0.032	0.009	0.006	-0.030
$Q(12)$	262.08	41.94	61.32	29.86	64.86
Squared Excess Returns					
$\hat{\rho}_1$	0.478	0.112	0.163	0.176	0.163
$\hat{\rho}_2$	0.281	0.149	0.155	0.087	0.161
$\hat{\rho}_3$	0.238	0.077	0.136	0.049	0.118
$\hat{\rho}_4$	0.290	0.020	0.111	0.087	0.173
$\hat{\rho}_5$	0.202	0.137	0.109	0.097	0.179
$Q(12)$	4543.79	404.14	560.52	286.52	527.91

$\hat{\rho}_\ell$ is the sample autocorrelation coefficient at lag ℓ with asymptotic standard error $1/\sqrt{T}$ and $Q(\ell)$ is the Box-Ljung portmanteau statistic based on ℓ squared autocorrelations.

Standard and Poor's Composite stock index series cover the period 1 January 1975 to 31 December 1998 whereas the Japanese Topix series starts on 1 January 1988 and ends at 31 December 1998. The stock data was obtained from Datastream. From the same data source we also collected daily UK and Japanese 1 month Treasury bill rates; the US 3 month Treasury bill rate data was extracted from the on-line Federal Reserve Bank of Chicago Statistical Release H.15 database. These interest rate series are used as proxies for the risk free rate of return. The stock index prices are in local currencies and not adjusted for dividends following studies of French, Schwert and Stambaugh (1987) and Poon and Taylor (1992) who found that inclusion of dividends affected estimation results only marginally. Returns are calculated on a continuously compounded basis and expressed in percentages, they are therefore calculated as $R_t = 100(\ln P_t - \ln P_{t-1})$ where P_t is the price of the stock market index at time t. From these returns we subtract the daily risk free rate multiplied by 100, denoted by Rf_t, in order to obtain the excess returns which are therefore defined as $y_t = R_t - Rf_t$.

In this section we model the behaviour of five series: we consider daily excess return series on the UK and US index that cover a period of 24 years ending in 1998, as well as 11 year sub-samples of these two series together with excess returns on the Japanese stock market index. These shorter series start in 1988 and therefore exclude the extreme negative observations relating to the 1987 stock market crash. Figures 3 and 4 contain graphs of the five excess return series, the accompanying summary

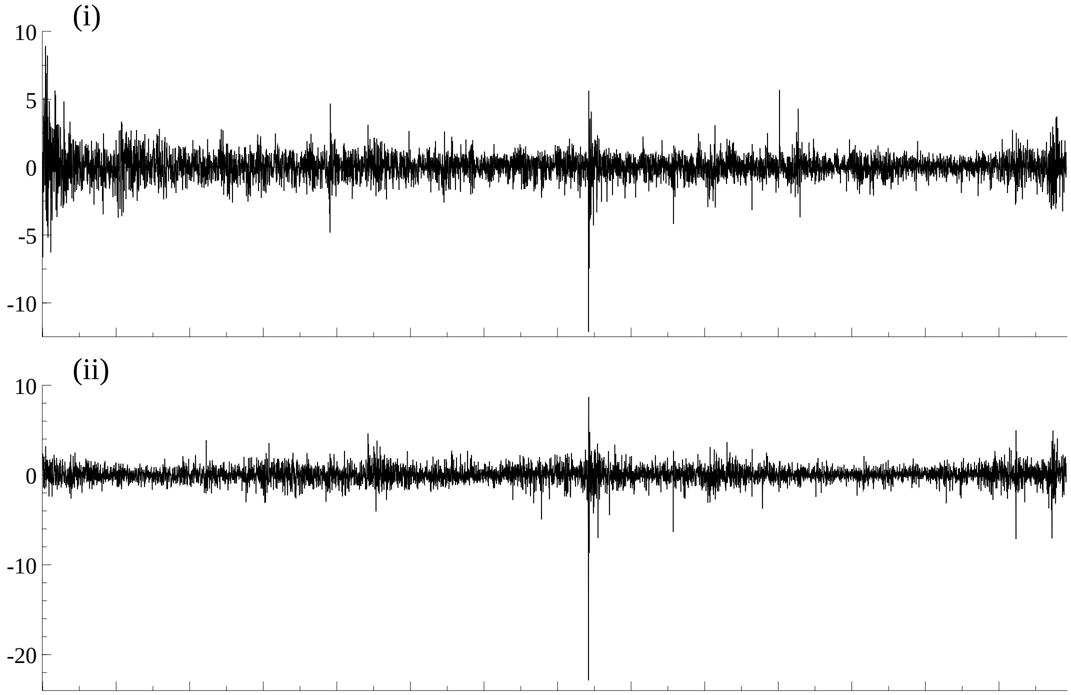

Figure 3: Excess returns for the (i) FT All Share Index (UK) and (ii) S&P Composite Stock Index (US) between 02/01/75 and 31/12/98

statistics are presented in table 1.

We observe that the effects of the October 1987 crash were especially pronounced for the US stock market where the Standard & Poor's Composite index fell by nearly 23% on one single trading day. This one observation contributed to a great extend to the large excess kurtosis value of 62.758 and the high negative skewness coefficient of -2.562. The most volatile series of the five is the Topix series which can not be attributed to one extreme movement, as can be seen in figure 4, but to several prolonged periods of market turbulence initiated in the early nineties by the collapse of the Japanese asset market. The Topix series is further characterised by a negative mean and is positively skewed, which are features not typically found in a stock index (excess) return series. We further observe that the UK excess returns and squared excess returns for the period starting in 1975 are highly autocorrelated at lag 1 but that these values are much lower and comparable with those of the Topix stock index for the sub-sample period 1988–1998. First-order serial correlation coefficients for the Standard & Poor's Composite Index excess returns on the other hand are relatively low for both the full and the sub-sample period. In the case of excess returns high first-order autocorrelation reflects the effects of non-synchronous or thin trading, whereas highly correlated squared returns can be seen as an indication of volatility clustering. The Q(12) test statistic, which is a joint test for the hypothesis that the first twelve autocorrelation coefficients are equal to zero, indicates that this hypothesis has to be rejected at the 1% significance level for all excess return and squared excess return series.

4.3 Estimation results for the SV(M) model and some diagnostics

Our main objective in this empirical section is to estimate the intertemporal relationship between excess returns on stock market indices and their volatility with our SVM model, which we already

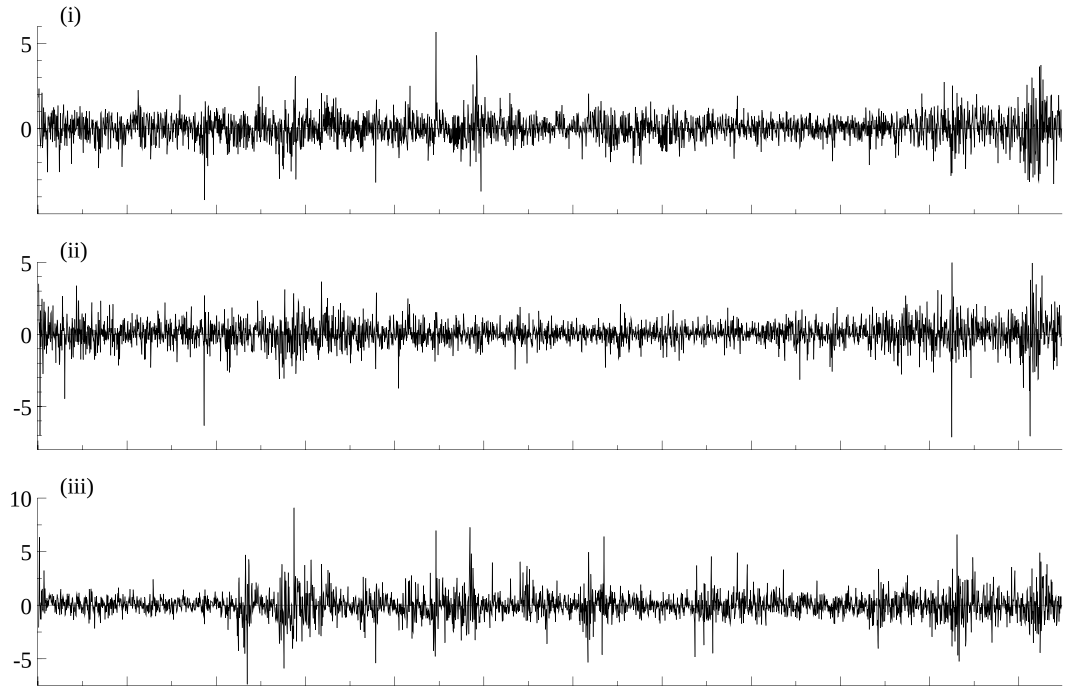

Figure 4: Excess returns for the (i) FT All Share Index (UK), (ii) S&P Composite Stock Index (US) and (iii) Topix Stock Index (Japan) between 04/01/88 and 31/12/98

defined in equations (5), (6) and (8).

In addition to this model we also estimate two alternative SV models which we obtain by imposing the constraints $d = 0$ and $a = b = d = 0$. With

$$y_t = \mu_t + \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0,1),$$

the mean equations of these models can be written as

SVM or SV-1 Model:	$\mu_t = a + b y_{t-1} + d\sigma^{*2} \exp(h_t)$
SV-2 Model:	$\mu_t = a + b y_{t-1}$
SV-3 Model:	$\mu_t = 0$

where all three SV models have the same variance specification

$$\begin{aligned} \sigma_t^2 &= \sigma^{*2} \exp(h_t), \\ h_t &= \phi h_{t-1} + \sigma_\eta \eta_t, \qquad \eta_t \sim \mathrm{NID}(0,1). \end{aligned}$$

In table 2 the estimation results for the SVM model are presented and our first observation is that the d parameter is negative for all five series. This implies that on average more volatile periods are associated with larger negative excess returns although the relationship is weak and the null hypothesis of d equal to zero can never be rejected at the conventional 5% level. The largest negative values for d are observed for the shorter period where they are also closest to being statistically significant. This

Table 2: Estimation results for the SVM (SV-1) model

Period	1975-1998		1988-1998		
No. of obs.	6261		2869		
Stock Index	FT All	S&P	FT All	S&P	Topix
a	0.038	0.045	0.061	0.074	0.019
	0.014 0.063	0.017 0.073	0.016 0.106	0.038 0.111	−0.014 0.052
b	0.146	0.074	0.100	0.024	0.099
	0.123 0.172	0.054 0.101	0.068 0.144	0.009 0.062	0.067 0.143
d	−0.011	−0.023	−0.085	−0.046	−0.031
	−0.049 0.028	−0.065 0.019	−0.176 0.006	−0.103 0.011	−0.066 0.005
σ^{*2}	0.615	0.597	0.458	0.539	0.832
	0.498 0.758	0.508 0.701	0.358 0.587	0.427 0.682	0.642 1.077
ϕ	0.984	0.979	0.976	0.970	0.966
	0.977 0.990	0.969 0.986	0.958 0.986	0.954 0.981	0.947 0.978
σ_η^2	0.018	0.021	0.019	0.035	0.058
	0.014 0.025	0.015 0.029	0.012 0.032	0.025 0.050	0.041 0.082
$\ln L$	-7531.80	-7496.80	-3039.36	-3299.35	-3988.34
AIC	15075.6	15005.6	6090.7	6610.7	7988.7
$Q(12)$	24.45	20.06	7.44	21.79	10.30
N	16.265	18.578	2.326	9.969	11.643
σ_∞^2	0.825	0.760	0.560	0.727	1.277
$corr(\varepsilon_t, \eta_{t+1})$	-0.033	-0.064	-0.070	-0.089	-0.175

Parameter estimates are reported together with the asymptotic 95% confidence interval which are a-symmetric for b, σ^{*2}, ϕ and σ_η^2; AIC is the Akaike Information Criterion which is calculated as -2(ln L) + 2p and $Q(\ell)$ is the Box-Ljung portmanteau statistic for the estimated observation errors which is asymptotically χ^2 distributed with $\ell - p$ degrees of freedom where p is the total number of estimated parameters; N is the χ^2 normality test statistic with 2 degrees of freedom; σ_∞^2 denotes the unconditional variance as implied by the volatility process.

is slightly surprising since we would expect the average association to become more positive once the effects of the 1987 stock market crash were no longer included in the sample. The relatively high parameter estimates for a indicate that the risk premium is not proportional to the variance of stock market returns. Except for the Topix series, which has a negative mean, the estimated values for a are positive and statistically significant. What is more, the estimate for the constant in the mean parameter consistently exceeds the mean value of the excess return series itself although this value is included in all confidence intervals for a, again with the exception of the Topix series. Estimates for the b parameter are all statistically significant and very similar to the first-order autocorrelation coefficients reported in table 1. The high first-order autocorrelation coefficient observed for the squared excess returns of the long FT All Share series is reflected in the persistence parameter estimate ϕ, which is close to one. The other four estimates for ϕ also lie in this region which is consistent with the near unity volatility persistence for high frequency data typically found with GARCH models. The more erratic behaviour of the Topix series is quite well captured by the SVM model through a combination of parameters: the scaling parameter σ^{*2} is quite large at 0.832 and the relatively small volatility persistence parameter ϕ, combined with a value of 0.058 for σ_η^2, implies that the Topix series is not only more volatile but also less predictable than any of the other four series.

With regard to the distributional assumptions we see that the standardised error term ε_t abides the normality assumption reasonably well, especially for the shorter UK series. This makes the need to specify alternative distributions for ε_t, such as the Student-t distribution which has fatter tails and is often employed in GARCH models, less imperative. The hypothesis that the first twelve autocorrelation coefficients of ε_t are equal to zero can not be rejected for the short FT All Share and the Topix series, as the critical value at the 5% significance level is 12.6. This indicates that there is little serial correlation left in the standardised error term.

We find that the correlation coefficients between the two error terms ε_t and η_{t+1} are consistently negative. This then implies that unexpected negative shocks to the excess returns are associated with increases in volatility, while unexpected positive shocks result in decreasing volatility values. We are inclined to interpret this as an indication of the presence of the leverage effect, or asymmetric volatility, even though our initial assumption was that of zero correlation between the two error terms.

Table 3: Estimation results for the SV-2 model

Period	1975-1998		1988-1998		
No. of obs.	6261		2869		
Stock Index	FT All	S&P	FT All	S&P	Topix
a	0.033	0.033	0.027	0.055	-0.0001
	0.016 0.050	0.017 0.049	0.004 0.050	0.032 0.078	-0.026 0.025
b	0.146	0.075	0.101	0.025	0.100
	0.122 0.174	0.054 0.102	0.069 0.146	0.010 0.060	0.088 0.113
σ^{*2}	0.615	0.597	0.459	0.540	0.831
	0.503 0.752	0.504 0.706	0.358 0.588	0.429 0.680	0.734 0.941
ϕ	0.984	0.979	0.976	0.970	0.965
	0.977 0.989	0.970 0.985	0.959 0.986	0.954 0.981	0.950 0.976
σ_η^2	0.018	0.020	0.019	0.035	0.059
	0.014 0.024	0.015 0.028	0.012 0.031	0.022 0.055	0.044 0.079
$\ln L$	-7531.92	-7496.36	-3040.80	-3300.30	-3989.32
$LR(d=0)$	0.25	1.13	2.87	1.89	1.95
AIC	15073.8	15002.7	6091.6	6610.6	7988.6
$Q(12)$	24.43	19.88	7.60	21.26	10.42
N	16.855	20.261	3.989	11.936	8.671
σ_∞^2	0.825	0.761	0.562	0.729	1.278
$corr(\varepsilon_t, \eta_{t+1})$	-0.034	-0.069	-0.081	-0.094	-0.179

Parameter estimates are reported together with the asymptotic 95% confidence interval which are a-symmetric for b, σ^{*2}, ϕ and σ_η^2; $LR(d=0)$ is the likelihood ratio statistic for the hypothesis $d=0$; AIC is the Akaike Information Criterion which is calculated as -2($\ln L$) + 2p and $Q(\ell)$ is the Box-Ljung portmanteau statistic for the estimated observation errors which is asymptotically χ^2 distributed with $\ell - p$ degrees of freedom where p is the total number of estimated parameters; N is the χ^2 normality test statistic with 2 degrees of freedom; σ_∞^2 denotes the unconditional variance as implied by the volatility process.

The results change only marginally when we estimate the SV-2 model where d is restricted to zero, results of which are presented in table 3. The main difference between the two models is the general decrease in the estimated value for a. The fact that the d parameter has little explanatory power is confirmed by the likelihood ratio test statistic which never exceeds the critical value χ_1^2 5% significance value of 3.84. All our parameter estimates are now statistically significant, with the exception of the a parameter for the Topix series. The AIC statistic, which is a goodness-of-fit statistic that allows comparison between models with different numbers of parameters, indicates that there is little to be gained by including the d parameter as this statistic favours the SV-2 model in four out of five cases.

Finally we present our findings for the SV-3 model in table 4 in order to compare the results with those of the SVM and the SV-2 model and determine whether SV models benefit from simultaneously

Table 4: Estimation results for the SV-3 model

Period	1975-1998		1988-1998		
No. of obs.	6261		2869		
Stock Index	FT All	S&P	FT All	S&P	Topix
σ^{*2}	0.631	0.602	0.465	0.546	0.839
	0.502 0.793	0.509 0.713	0.365 0.592	0.424 0.703	0.654 1.076
ϕ	0.985	0.979	0.976	0.971	0.965
	0.980 0.989	0.971 0.985	0.962 0.985	0.948 0.984	0.946 0.977
σ_η^2	0.017	0.020	0.019	0.033	0.059
	0.014 0.022	0.014 0.027	0.012 0.031	0.019 0.058	0.041 0.085
$\ln L$	-7604.05	-7521.84	-3057.78	-3311.34	-4002.63
$LR(a=b=d=0)$	144.5	50.1	36.8	24.0	28.6
AIC	15214.1	15049.7	6121.6	6628.7	8011.3
$Q(12)$	151.76	55.40	35.49	23.02	39.77
N	18.422	31.913	4.503	18.788	4.914
σ_∞^2	0.849	0.765	0.570	0.732	1.288
$corr(\varepsilon_t, \eta_{t+1})$	-0.055	-0.086	-0.101	-0.123	-0.186

Parameter estimates are reported together with the asymptotic 95% confidence interval which are a-symmetric for σ^{*2}, ϕ and σ_η^2; $LR(a=b=d=0)$ is the likelihood ratio statistic for the hypothesis $a=b=d=0$; AIC is the Akaike Information Criterion which is calculated as -2($\ln L$) + 2p and $Q(\ell)$ is the Box-Ljung portmanteau statistic for the estimated observation errors which is asymptotically χ^2 distributed with $\ell - p$ degrees of freedom where p is the total number of estimated parameters; N is the χ^2 normality test statistic with 2 degrees of freedom; σ_∞^2 denotes the unconditional variance.

modelling of both the mean and the variance equation or not. On the basis of the likelihood ratio test statistics the conclusion would have to be that simultaneous estimation is quite advantageous as these values are always statistically significant at the 1% confidence level, which is confirmed by the values for the AIC statistic. We further note that ε_t is in general less well-behaved for the SV-3 than for the other two Stochastic Volatility models, especially in terms of the assumption of zero autocorrelation.

With regard to the Stochastic Volatility in Mean model we can conclude that our findings are different from those usually observed in the GARCH literature where relatively small values for a and positive estimates for the in-mean parameter have been reported for both the UK and the US stock market, although there is also evidence of an average negative relationship between excess returns and volatility for the US market. In order to compare both methods in more detail we present GARCH estimates for our five series in the next subsection.

4.4 Some comparisons with GARCH estimation results

Our initial attention will be on the estimation results of the GARCH in Mean model as defined in equations (4) and (9). With $y_t = \mu_t + \sigma_t \varepsilon_t$, its mean equation is therefore expressed as

$$\textit{GARCH-M or GA-1 Model:} \qquad \mu_t = a + b y_{t-1} + d\sigma_t^2$$

The GA-2 and GA-3 model have mean specifications identical to those of the SV-2 and SV-3 model, respectively, and all three GARCH models have their variance defined by

$$\sigma_t^2 = \omega + \alpha(y_{t-1} - \mu_{t-1})^2 + \beta\sigma_{t-1}^2.$$

Estimation results for the GARCH-M or GA-1 model are given in table 5 where we observe that near zero estimates for a are combined with positive values for d. The null hypothesis of a zero

Table 5: Estimation results for the GARCH-M (GA-1) model

Period	1975-1998		1988-1998		
No. of obs.	6261		2869		
Stock Index	FT All	S&P	FT All	S&P	Topix
a	0.022	0.003	−0.010	0.004	−0.015
	−0.008 0.052	−0.034 0.039	−0.063 0.043	−0.050 0.059	−0.063 0.033
b	0.152	0.080	0.115	0.019	0.112
	0.127 0.178	0.052 0.107	0.076 0.154	−0.019 0.057	0.072 0.151
d	0.017	0.046	0.070	0.061	0.043
	−0.024 0.058	−0.005 0.099	−0.033 0.172	−0.023 0.145	−0.002 0.087
ω	0.018	0.011	0.014	0.003	0.032
	0.014 0.023	0.008 0.014	0.008 0.020	0.002 0.005	0.023 0.040
α	0.094	0.065	0.072	0.032	0.134
	0.082 0.105	0.062 0.068	0.057 0.087	0.027 0.037	0.116 0.152
β	0.885	0.924	0.905	0.965	0.849
	0.871 0.899	0.918 0.930	0.882 0.927	0.959 0.971	0.829 0.869
$\alpha + \beta$	0.978	0.989	0.976	0.997	0.983
$\ln L$	-7615.05	-7667.26	-3078.86	-3416.20	-4078.56
AIC	15242.1	15346.5	6169.7	6844.4	8169.1
$Q(12)$	35.81	11.89	10.16	17.13	11.25
N	6630.1	10254.1	390.7	6022.5	810.8
σ^2_∞	0.852	0.973	0.583	0.991	1.823

Parameter estimates are reported together with the asymptotic 95% confidence interval which are all symmetric; AIC is the Akaike Information Criterion which is calculated as -2($\ln L$) + 2p and $Q(\ell)$ is the Box-Ljung portmanteau statistic for the estimated observation errors which is asymptotically χ^2 distributed with $\ell - p$ degrees of freedom where p is the total number of estimated parameters; N is the χ^2 normality test statistic with 2 degrees of freedom; σ^2_∞ denotes the unconditional variance as implied by the volatility process.

relationship between excess returns and volatility can however never be rejected at the 5% significance level, although in some cases only by a very small margin. The main difference between the SVM and GARCH-M model is therefore that the parameter estimates for d are negative for the SVM model and positive for the GARCH-M model, although none of these are ever statistically significant. Further, a is statistically significant in the SVM model and insignificant in the GARCH-M model. The values for a in the SVM model are relatively large and this is of course a natural consequence of the negative parameter estimate for d. A very similar pattern has been observed before in the literature by Glosten et al. (1993) who developed a GARCH-M model which included a number of additional variables in both the mean and variance equation. For the standard GARCH-M model (which is identical to our GA-1 model but with b constrained to zero) they initially found a small parameter estimate for a and a positive value for d. Re-estimation with their extended GARCH-M model resulted however in much larger values for the constant in the mean combined with negative values for the in-mean parameter which, on occasion, were even statistically significant. Our SVM model therefore appears to be closer related to this richer GARCH-M model than to the GA-1 model we estimated here.

In section 4.1 we discussed why an average negative intertemporal relationship between excess returns on a stock index and volatility might be more likely than a positive one. This does however not explain why the two in-mean models result in such different estimates for the d parameter. One of the reasons might be the different definitions of the two volatility processes. The SVM model immediately incorporates the effect of an unexpected return shock in the volatility process through η_t, whereas the GARCH-M model does not absorp this new information until time $t+1$. Following an unanticipated shock at time t the variance of the GARCH model only starts to increase at time $t+1$ and it is not until the subsequent period that it becomes fully incorporated. Consequently the d parameter in the GARCH-M model does not actually measure the contemporaneous relationship between expected returns and volatility. This problem is most pronounced at the beginning of a volatile period when the shock to the return process is large and the GARCH-M variance still small. The relevant question therefore appears to be how many volatile periods in the return series start with a large negative rather than a positive shock, as this explains the difference between the sign of the d parameter for the two models. As volatile periods are usually initiated by large negative unexpected returns, the d parameter is bound to be larger for the GARCH-M than for the SVM model. What is more, negative unanticipated returns also induce more volatile behaviour than positive ones as shown in many empirical applications of Nelson's EGARCH model and affirmed by the negative correlation between the two estimated error terms ε_t and η_{t+1} we observed in our SV model[8].

Although the size of the shock is certainly relevant, it does not seem to be of crucial importance as the d parameter in the SVM model is smaller for the longer Financial Times All Share and Standard & Poor's Composite Index series than for the same series starting in 1988.

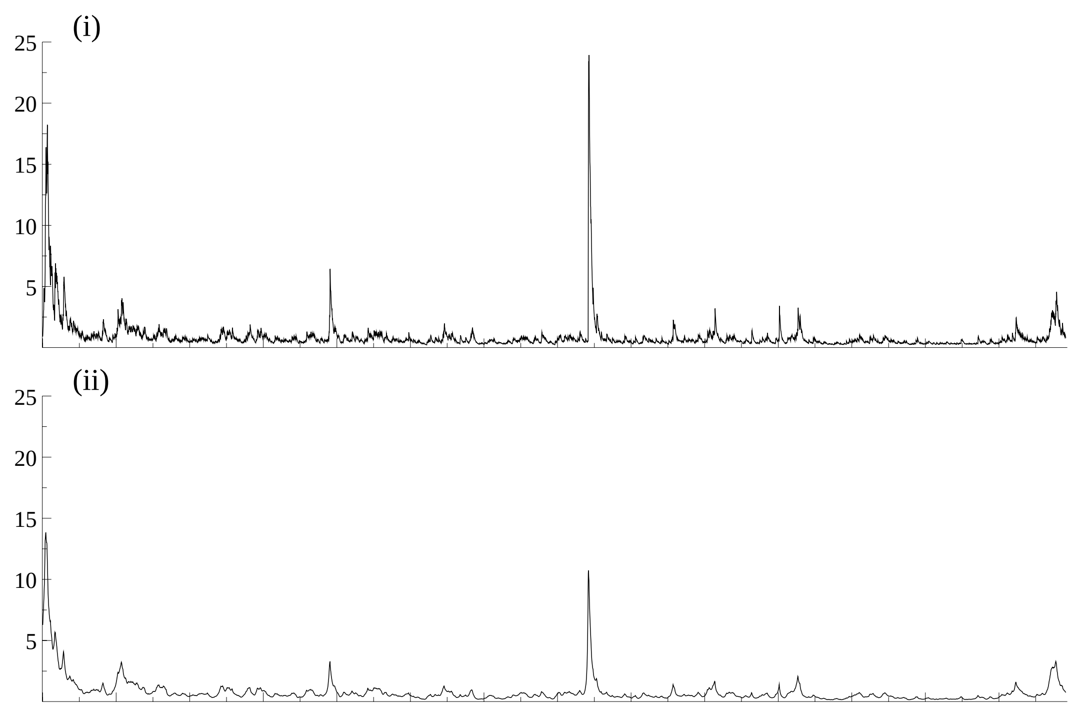

Figure 5: Estimated volatility for FT All Share Index (UK) between 03/01/75 and 31/12/98 as implied by the (i) GARCH-M model and (ii) SVM model

Another difference between the two volatility series becomes evident when we graph the variance implied by the SVM and the GARCH-M model for the Financial Times All Share Index in figure 5 for the period starting in 1975. The variance for the SVM model is given by $\sigma^{*2}\exp(h_t)|I_T$ and that of the GARCH-M model by $\sigma_t^2|I_{t-1}$ where T denotes the sample size. The graph clearly shows that the

[8] Also see the discussion and references in Harvey and Shephard (1996)

GARCH-M variance series reacts more abruptly than the variance series implied by the SVM model. The main reason for this is that the SVM model produces a smoothed volatility series based on all information in the sample (I_T), whereas the GARCH-M (1,1) model is a conditional model based on information available at time $t-1$. The filtered estimate of the volatility for the SVM model, that is $\sigma^{*2}\exp(h_t)|I_{t-1}$ can be obtained via the technique of particle filtering, see Pitt and Shephard (1999). We further note that both series broadly follow the same pattern and are closely related with correlation 0.86, although the average value for the GARCH model is considerably higher than that of the SVM model: 0.90 for the GARCH-M$(1,1)$ series as opposed to an average value of 0.79 when the volatility of the Financial Times All Share Index is estimated with the SVM model[9].

Table 6: Estimation results for the GA-2 model

Period	1975-1998		1988-1998		
No. of obs.	6261		2869		
Stock Index	FT All	S&P	FT All	S&P	Topix
a	0.032	0.030	0.022	0.038	0.021
	0.013 0.050	0.011 0.049	−0.003 0.047	0.010 0.066	−0.011 0.052
b	0.152	0.080	0.116	0.020	0.113
	0.127 0.178	0.053 0.107	0.077 0.155	−0.018 0.057	0.075 0.151
ω	0.018	0.011	0.014	0.003	0.031
	0.014 0.023	0.008 0.013	0.008 0.020	0.002 0.005	0.023 0.039
α	0.094	0.065	0.072	0.031	0.134
	0.082 0.105	0.062 0.067	0.057 0.087	0.026 0.037	0.116 0.152
β	0.885	0.924	0.904	0.965	0.850
	0.871 0.899	0.918 0.930	0.881 0.927	0.959 0.971	0.830 0.869
$\alpha+\beta$	0.978	0.989	0.976	0.997	0.984
$\ln L$	-7615.39	-7669.95	-3079.79	-3417.49	-4080.41
$LR(d=0)$	0.68	5.38	1.86	2.58	3.70
AIC	15240.8	15349.9	6169.6	6845.0	8170.8
$Q(12)$	35.65	12.36	9.57	17.79	9.78
N	6600.8	10478.1	392.2	5979.3	811.4
σ^2_∞	0.852	0.978	0.584	1.002	1.884

Parameter estimates are reported together with the asymptotic 95% confidence interval which are all symmetric; $LR(d=0)$ is the likelihood ratio statistic for the hypothesis $d=0$; AIC is the Akaike Information Criterion which is calculated as -2($\ln L$) + 2p and $Q(\ell)$ is the Box-Ljung portmanteau statistic for the estimated observation errors which is asymptotically χ^2 distributed with $\ell-p$ degrees of freedom where p is the total number of estimated parameters; N is the χ^2 normality test statistic with 2 degrees of freedom; σ^2_∞ denotes the unconditional variance as implied by the volatility process.

Estimation results for the remaining two GARCH specifications, the GA-2 and the GA-3 model are presented in tables 6 and 7. Comparison between SV and GARCH models shows that estimates for b are very similar across the various model specifications. The volatility persistence parameters are all close to unity although we find that the persistence values for the Topix and the Standard & Poor's series starting in 1988 are considerably higher when the volatility process is modelled with GARCH models. They are in fact so high that they exceed those of the 1975–1998 Financial Times All Share

[9]Variance series of this index have been examined previously in the literature: Poon and Taylor (1992) graphed and compared a monthly GARCH(1,1) conditional variance series with an ARMA$(1,1)$ variance series for the period 1969–1989. They also found larger average values for the GARCH series and a correlation coefficient between the two variance series of 0.87.

Table 7: Estimation results for the GA-3 model

Period	1975-1998		1988-1998		
No. of obs.	6261		2869		
Stock Index	FT All	S&P	FT All	S&P	Topix
ω	0.018	0.011	0.014	0.004	0.033
	0.014 0.022	0.008 0.013	0.008 0.020	0.002 0.006	0.025 0.041
α	0.093	0.064	0.074	0.036	0.134
	0.082 0.104	0.061 0.066	0.060 0.089	0.031 0.041	0.117 0.151
β	0.887	0.926	0.902	0.960	0.848
	0.874 0.901	0.920 0.931	0.880 0.925	0.954 0.966	0.829 0.868
$\alpha + \beta$	0.980	0.995	0.976	0.996	0.983
$\ln L$	-7707.40	-7696.96	-3101.58	-3421.38	-4096.63
$LR(a = b = d = 0)$	184.7	59.4	45.4	10.4	36.1
AIC	15420.8	15399.9	6209.2	6848.8	8199.3
$Q(12)$	207.02	61.47	48.76	21.36	47.06
N	7615.5	10192.5	429.6	4497.2	825.7
σ^2_∞	0.905	0.995	0.599	0.964	1.857

Parameter estimates are reported together with the asymptotic 95% confidence interval which are all symmetric; $LR(a = b = d = 0)$ is the likelihood ratio statistic for the hypothesis $a = b = d = 0$; AIC is the Akaike Information Criterion which is calculated as -2($\ln L$) + 2p and $Q(\ell)$ is the Box-Ljung portmanteau statistic for the estimated observation errors which is asymptotically χ^2 distributed with $\ell - p$ degrees of freedom where p is the total number of estimated parameters; N is the χ^2 normality test statistic with 2 degrees of freedom; σ^2_∞ denotes the unconditional variance.

Index series which exhibits very high autocorrelated squared returns as shown in table 1. We further observe that unconditional variances are consistently lower for SV models and that diagnostic statistics with regard to the standardised residual ε_t seem to favour SV models in most cases, especially in terms of the normality test statistic.

5 Summary and Conclusions

In this paper we have presented a Stochastic Volatility model where the mean is modelled simultaneously with the variance equation. When one of the variables in the mean is the volatility process itself, we obtain the Stochastic Volatility in Mean (SVM) model with which we are able to investigate the contemporaneous relationship between expected excess returns on a stock market index and its time-varying volatility. We estimate the parameters in our model using a special simulation based maximum likelihood method and we also present results of a simulation experiment to show that if such a interdependence is present our SVM model is capable of detecting it.

For our empirical application we examined stock indices from the United Kingdom, the United States and Japan over two time periods and for three different mean equations. The results were then compared with the estimation results obtained for their GARCH counterparts. Our conclusions can be summarised as follows. Firstly, with our SVM model we find evidence of a weak negative relationship for all stock index series, whereas estimation with the GARCH-M model produces positive, but again statistically insignificant, estimates for the in-mean parameter d. We assert however that a negative average relationship between excess returns and their contemporaneous time-varying volatility is more plausible than a positive one and that the sign of the in-mean parameter can be at least partially explained by the difference in definitions of the volatility process of the models. The large positive value for the constant in the mean parameter a observed in the SVM model, as opposed to a near-zero estimate in the case of the GARCH-M model, is a natural consequence of the negative value for d. The first-order autoregressive term b in the mean equation appears robust across model specifications and classes of volatility models. Secondly, we find that simultaneous modelling of the mean and the variance equation results in improvements in terms of the goodness-of-fit of the model. Although it is possible to model the original series prior to estimation with a volatility model, simultaneous estimation is more efficient. Finally, we observe that the volatility persistence parameter ϕ in the SV models, which is an indication of volatility clustering, is comparable with those of GARCH models and might even be preferable to the latter. An additional advantage of SV models over GARCH models is that the distributional assumptions of the error term in the mean ε_t are much less violated by our SV model, especially in terms of the normality assumption. This makes the case for departures from normality and hence the estimation of additional parameters less strong. On the basis of the above we therefore feel that SV models can be regarded as a more than competitive alternative to GARCH models, not only in theoretical terms but also in empirical research.

Appendix: approximating model used for simulation

The approximating model is based on a linear Gaussian model with mean $\mathrm{E}(y_t) = h_t + c_t$ and variance $\mathrm{V}(y_t) = H_t$, that is

$$y_t = h_t + u_t, \qquad u_t \sim \mathrm{N}(c_t, H_t), \qquad t = 1, \ldots, n, \tag{19}$$

where c_t and H_t are determined in such a way that the mean and variance of y_t implied by the approximating model (19) and by the true model (1) and (7) are as close as possible[10].

We achieve this by equalising the first and second derivatives of $p(y|\theta, \psi)$ and $g(y|\theta, \psi)$ with respect to θ at $\hat{\theta} = \tilde{\mathrm{E}}(\theta) = \int \theta g(\theta|y, \psi)$. Note that $p(\cdot)$ refers to a density for the true model and $g(\cdot)$ refers to a density for the approximating Gaussian model. Further, it follows that $\hat{\theta}$ can simply be obtained via the Kalman filter and smoother applied to the approximating model (19). The conditional densities are given by

$$p(y|\theta, \psi) = \prod_{t=1}^{n} p_t, \qquad g(y|\theta, \psi) = \prod_{t=1}^{n} g_t,$$

with

$$\begin{aligned} p_t &= p(y_t|h_t, \psi) = -0.5[\log 2\pi\sigma^2 + h_t + \exp(-h_t)\sigma^{-2}\{y_t - d\exp(h_t)\}^2], \\ g_t &= g(y_t|h_t, \psi) = -0.5\{\log 2\pi + \log H_t + H_t^{-1}(y_t - c_t - h_t)^2\}. \end{aligned}$$

Differentiating both densities twice with respect to h_t gives

$$\begin{aligned} \dot{p}_t &= -0.5[1 + \sigma^{-2}\{d^2\sigma^{*2}\exp(h_t) - y_t^2\exp(-h_t)\}], \\ \ddot{p}_t &= -0.5\sigma^{-2}[d^2\sigma^{*2}\exp(h_t) + y_t^2\exp(-h_t)], \\ \dot{g}_t &= H_t^{-1}(y_t - c_t - h_t), \\ \ddot{g}_t &= -H_t^{-1}. \end{aligned}$$

Equalising the first and second derivatives, that is $\dot{p}_t = \dot{g}_t$ and $\ddot{p}_t = \ddot{g}_t$ for $t = 1, \ldots, n$, leads to

$$\begin{aligned} c_t &= y_t - h_t + 0.5H_t[1 + \sigma^{-2}\{d^2\sigma^{*2}\exp(h_t) - y_t^2\exp(-h_t)\}], \\ H_t &= 2\sigma^{*2}/[d^2\sigma^{*2}\exp(h_t) + y_t^2\sigma^{*-2}\exp(-h_t)]. \end{aligned}$$

[10]Note that the true model implies a nonlinear relationship between y_t and h_t; the approximating (linear) model is effectively a second-order Taylor expansion of the true model around h_t.

The resulting model for $\tilde{y}_t = y_t - c_t$ is equivalent to

$$\tilde{y}_t = h_t + \tilde{u}_t, \qquad \tilde{u}_t \sim \mathrm{N}(0, H_t), \qquad t = 1, \dots, n,$$

with

$$\tilde{y}_t = h_t - \frac{\sigma^{*2} + d^2\sigma^{*2}\exp(h_t) - y_t^2\sigma^{*-2}\exp(-h_t)}{d^2\sigma^{*2}\exp(h_t) + y_t^2\sigma^{*-2}\exp(-h_t)}, \qquad H_t = \frac{2\sigma^2}{d^2\sigma^{*2}\exp(h_t) + y_t^2\sigma^{*-2}\exp(-h_t)}.$$

It should be noted that $H_t > 0$ for any value of h_t. We cannot solve out for $\tilde{y}_t$ and H_t at $\hat{h}_t = \tilde{\mathrm{E}}(h_t)$ because $\tilde{\mathrm{E}}$ refers to expectation with respect to the approximating model which depend on h_t. However, such complicated but linear system of equations is usually solved iteratively by starting with a trial value $h_t = h_t^*$. Computing $\tilde{y}_t$ and H_t based on h_t^* and applying the Kalman filter smoother to model (19) leads to a smoothed estimate for h_t which can be used as a new trial value for h_t. Recomputing $\tilde{y}_t$ and H_t based on this new trial value leads to an iterative procedure which converges to $\hat{h}_t$. Note that the first and second derivatives of the true and approximating densities are equal at $h_t = \hat{h}_t$. More details are given by Durbin and Koopman (1997). It is worth mentioning that $\hat{h}_t$ is equal to the mode of $p(h_t|y, \psi)$ which can be of interest.

Acknowledgements

Most part of this work was conducted while the first author was at CentER of Tilburg University and the second author was visiting CentER. In particular, we like to thank Theo Nijman for his support in this project. Further, we are indebted to the participants of the Financial Markets Group meeting "Limits to Forecasting in Financial Markets" at the London School of Economics in June 1999 for helpful comments and suggestions. Also, for most part of this project, the first author was a Research Fellow of the Royal Netherlands Academy of Arts and Sciences and its financial support is gratefully acknowledged. The second author would especially like to thank John Cadle (Birmingham University).

References

Bera, A.K. and M.L. Higgins (1993), ARCH Models: Properties, Estimation and Testing, *Journal of Economic Surveys* 7, 305-366.

Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31, 307-327.

Bollerslev, T., R.F. Engle and D.B. Nelson (1994), ARCH Models, in: *Handbook of Econometrics*, Vol. 4, eds. R.F. Engle and D.L. McFadden, Elsevier Science, Amsterdam, 2959-3038.

Bollerslev, T., R.Y. Chou and K.F. Kroner (1992), ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence, *Journal of Econometrics* 52, 5-59.

Campbell, J.Y. and L. Hentschel (1992), No News is Good News, An Asymmetric Model of Changing Volatility in Stock Returns, *Journal of Financial Economics* 31, 281-318.

Campbell, J.Y., A.W. Lo and A.C. MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.

Chesney, M. and L. Scott (1989). Pricing European Currency Options: A Comparison of the Modified Black-Scholes Model and a Random Variance Model, *Journal of Financial and Quantitative Analysis* 24, 267-284.

Chib, S., F. Nardari and N. Shephard (1998), Markov Chain Monte Carlo Methods for Generalized Stochastic Volatility Models, Discussion Paper: Nuffield College, Oxford.

de Jong, P. and N. Shephard (1995), The Simulation Smoother for Time Series Models, *Biometrika* 82, 339-350.

Danielsson, J. (1994), Stochastic Volatility in Asset Prices, Estimation with Simulated Maximum Likelihood, *Journal of Econometrics* 64, 375-400.

Diebold, F.X. and J.A. Lopez (1995), Modeling Volatility Dynamics, in: *Macroeconometrics: Developments, Tensions and Prospects*, eds. K. Hoover, Kluwer Academic Publishers, Amsterdam, 427-472.

Doornik, J.A. (1998), *Object-Oriented Matrix Programming using Ox 2.0*, Timberlake Consultants Press, London.

Durbin, J. and S.J. Koopman (1997), Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models, *Biometrika* 84, 669-84.

Durbin, J. and S.J. Koopman (2000), Time Series Analysis of Non-Gaussian Observations based on State Space Models from both Classical and Bayesian Perspectives (with discussion), *Journal of the Royal Statistical Society, Series B*, forthcoming.

Engle, R.F., D.M. Lilien and R.P. Robins (1987), Estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M Model, *Econometrica* 55, 391-407.

French, K.R., G.W. Schwert and R.F. Stambaugh (1987), Expected Stock Returns and Volatility, *Journal of Financial Economics* 19, 3-29.

Fridman, M. and L. Harris (1998), A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models, *Journal of Business and Economic Statistics* 16, 284-291.

Gallant, A.R., P.E. Rossi and G. Tauchen (1992), Stock Prices and Volume, *Review of Financial Studies* 5, 199-242.

Ghysels, E., A.C. Harvey and E. Renault (1996), Stochastic Volatility, in: *Handbook of Statistics*, Vol. 14, Statistical Methods in Finance, eds. G.S. Maddala and C.R. Rao, North-Holland, Amsterdam, 128-198.

Glosten, L.R., R. Jagannathan and D.E. Runkle (1993), On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance* 48, 1779-1801.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A.C., E. Ruiz and N. Shephard (1994), Multivariate Stochastic Variance Models, *Review of Economic Studies* 61, 247-264.

Harvey, A.C. and N. Shephard (1996), Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns, *Journal of Business and Economic Statistics* 14, 429-434.

Hull, J. and A. White (1987), The Pricing of Options on Assets with Stochastic Volatilities, *The*

Journal of Finance 42, 281-300.

Jacquier, E., N.G. Polson and P.E. Rossi (1994), Bayesian Analysis of Stochastic Volatility Models (with discussion), *Journal of Business and Economic Statistics* 12, 371- 417.

Kim, S., N. Shephard and S. Chib (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies* 65, 361-393.

Koopman, S.J., Harvey, A.C., Doornik, J.A., and Shephard, N. (1995), *STAMP 5.0 Structural Time Series Analyser, Modeller and Predictor*, Chapman and Hall, London.

Koopman, S.J., N. Shephard and J. Doornik (1999), Statistical Algorithms for Models in State Space using SsfPack 2.2, *Econometrics Journal*, 2, 113-166.

Nelson, D.B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* 59, 347-370.

Pitt, M.K. and N. Shephard (1999), Filtering via simulation: auxiliary particle filter, *Journal of the American Statistical Association*, 94, 590-599.

Poon, S. and S.J. Taylor (1992), Stock Returns and Volatility: An Empirical Study of the UK Stock Market, *Journal of Banking and Finance* 16, 37-59.

Ripley, B.D. (1987), *Stochastic Simulation*, Wiley, New York.

Sandmann, G. and S.J. Koopman (1998), Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood, *Journal of Econometrics* 87, 271-301.

Scott, L.O. (1987), Option Pricing when the Variance Changes Randomly: Theory, Estimation and Application, *Journal of Financial and Quantitative Analysis* 22, 419-438.

Shephard, N.G. (1996), Statistical Aspects of ARCH and Stochastic Volatility, in: *Time Series Models in Econometrics, Finance and Other Fields*, Monographs on Statistics and Applied Probability 65, eds. D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, Chapman and Hall, London, 1-67.

Shephard, N. and M.K. Pitt (1997), Likelihood Analysis of Non-Gaussian Measurement Time Series. *Biometrika* 84, 653-667.

Taylor, S.J. (1986), *Modelling Financial Time Series*, John Wiley and Sons, Chichester.

Taylor, S.J. (1994), Modeling Stochastic Volatility: A Review and Comparative Study, *Mathematical Finance* 4, 183-204.

Wiggins, J.B. (1987), Option Values under Stochastic Volatility: Theory and Empirical Estimates, *Journal of Financial Economics* 19, 351-372.